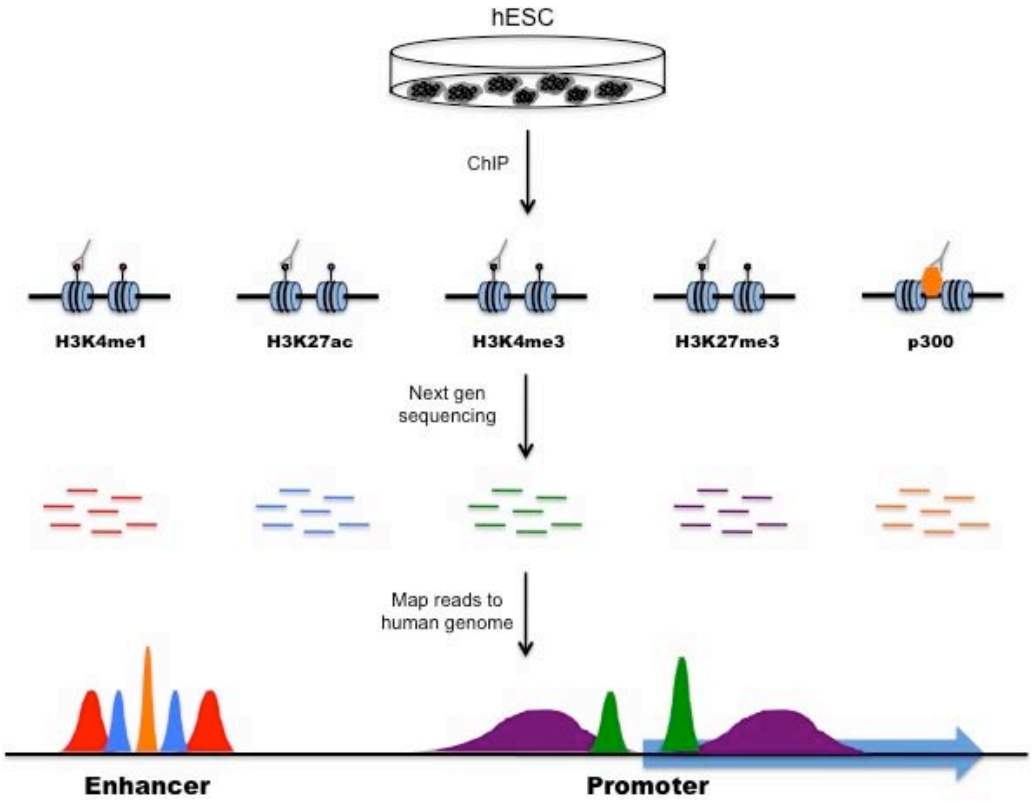
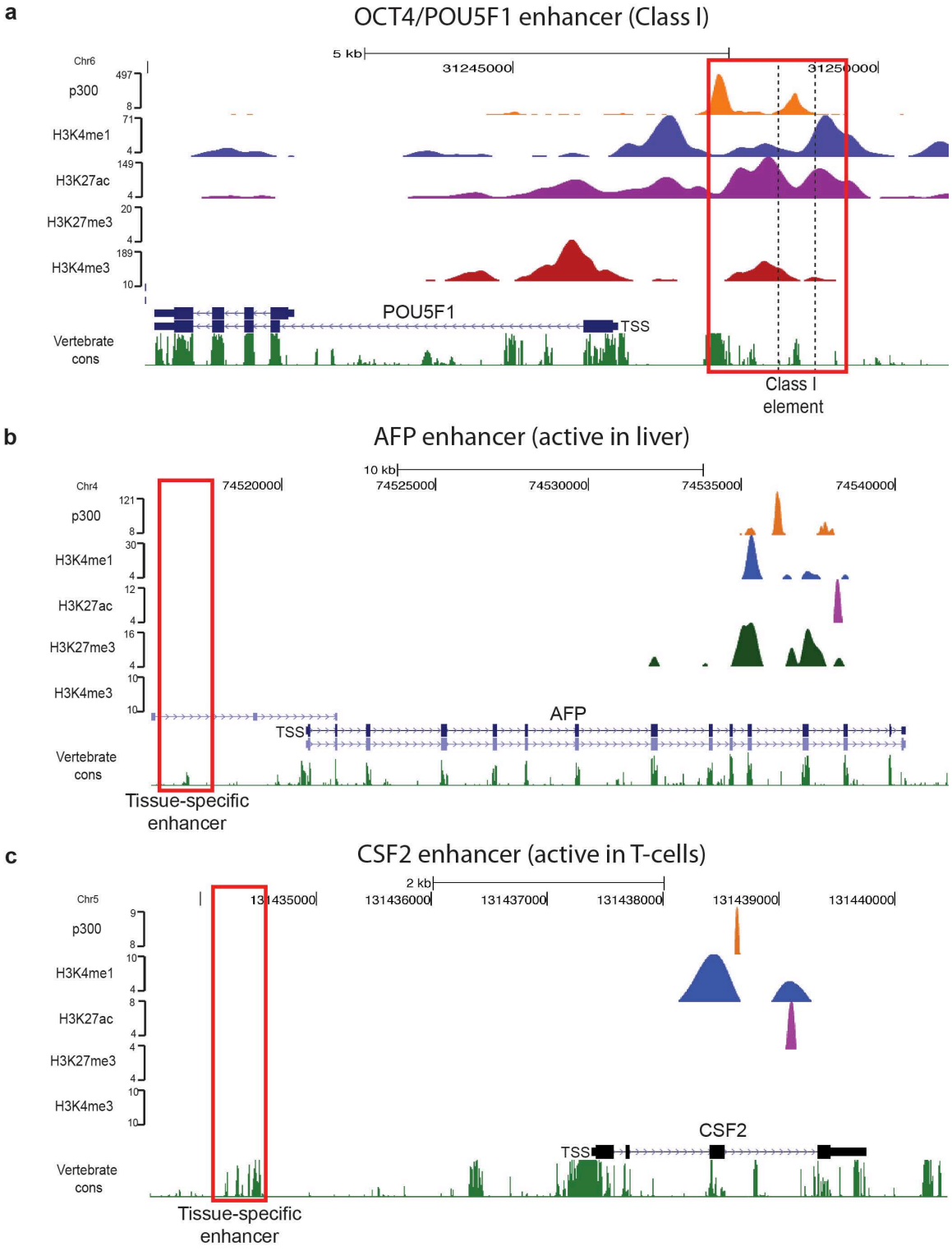
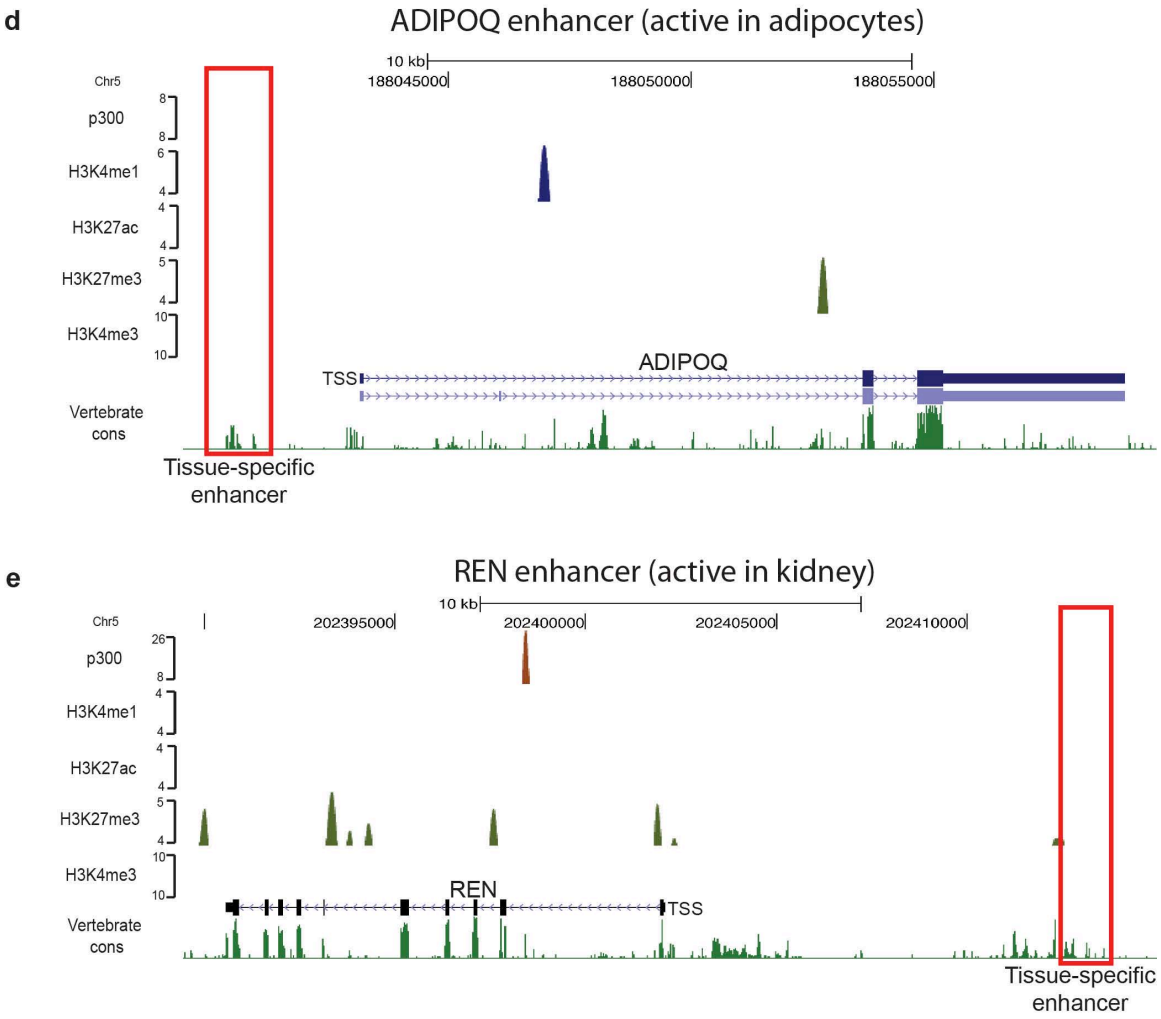
doi:10.1038/nature09692

SUPPLEMENTARY FIGURES AND TABLES

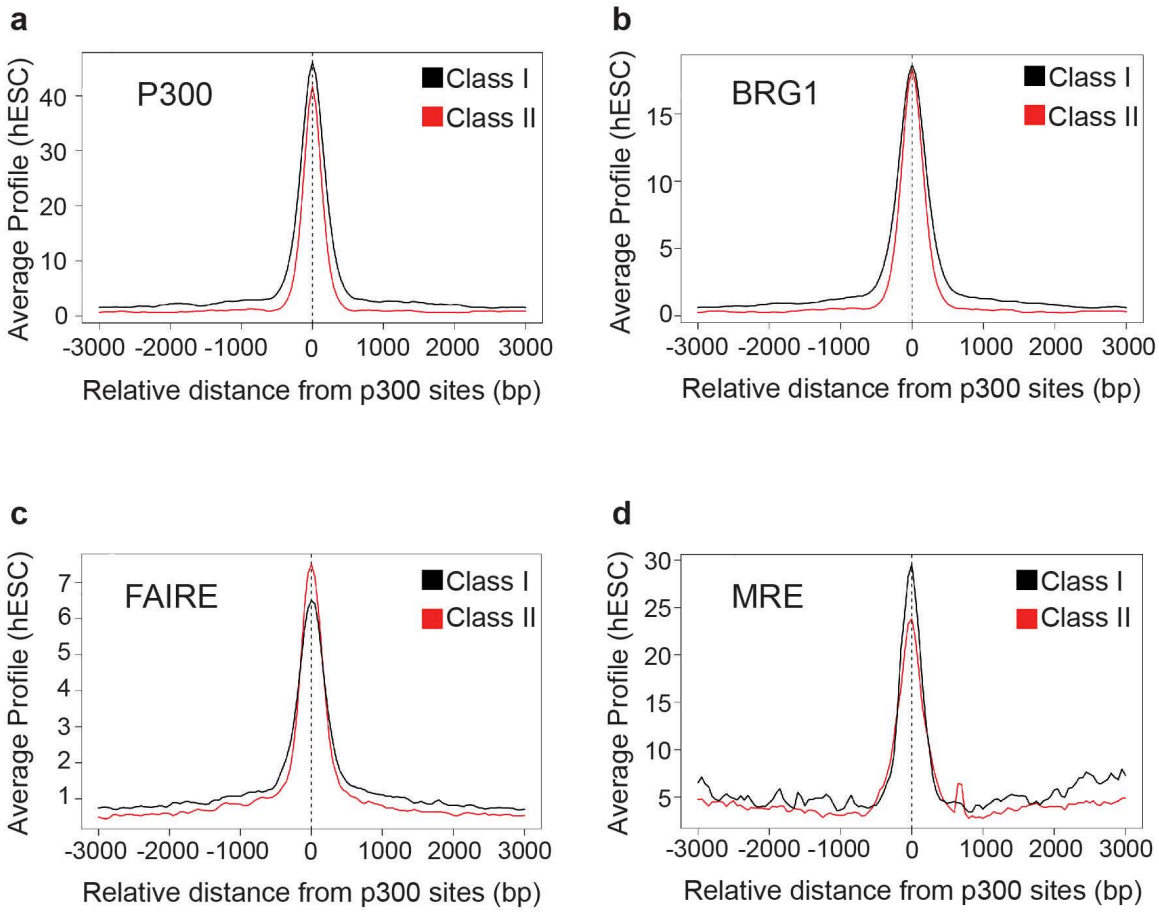


Supplementary Figure 1. Strategy for the identification of the hESC enhancer repertoire. Sonicated hESC chromatin was used for ChIPs with antibodies against p300, H3K4me1 and H3K27ac, which have been previously shown to mark active enhancer elements, and with antibodies against H3K4me3 and H3K27me3, which have been previously shown to mark proximal promoter regions. The resulting DNA fragments were then sequenced, along with total genomic DNA (input). Sequencing reads were mapped to the human genome, and sequence-tag enrichment regions were determined using QuEST.


a
OCT4/POU5F1 enhancer (Class I)
Chr6
5 kb
31245000
31250000
p300
H3K4me1
H3K27ac
H3K27me3
H3K4me3
POU5F1
TSS
Vertebrate cons
Class I element
b
AFP enhancer (active in liver)
Chr4
10 kb
74520000
74525000
74530000
74535000
74540000
p300
H3K4me1
H3K27ac
H3K27me3
H3K4me3
AFP
TSS
Vertebrate cons
Tissue-specific enhancer
c
CSF2 enhancer (active in T-cells)
Chr5
2 kb
131435000
131436000
131437000
131438000
131439000
131440000
p300
H3K4me1
H3K27ac
H3K27me3
H3K4me3
CSF2
TSS
Vertebrate cons
Tissue-specific enhancer

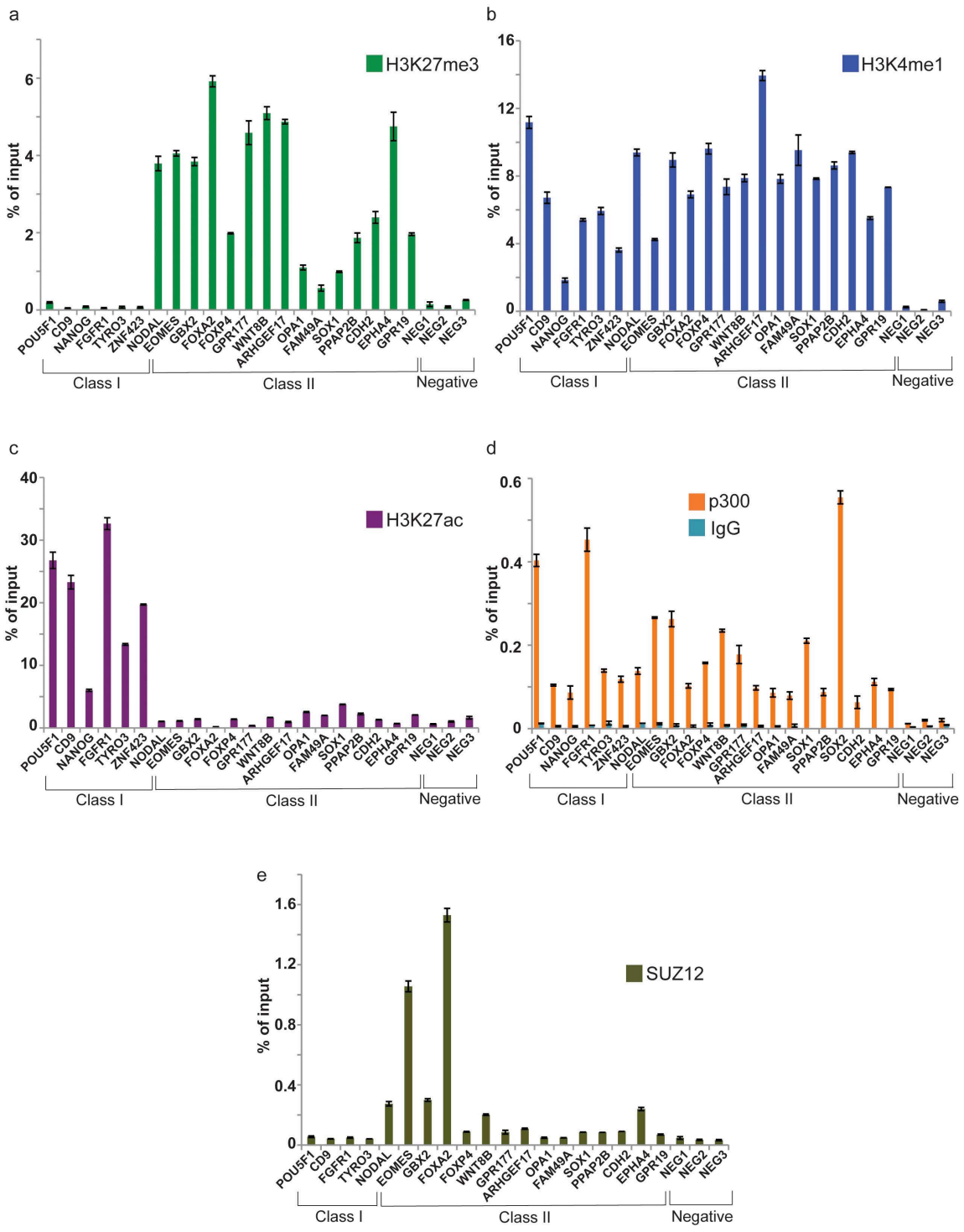


Supplementary Figure 2. Genome browser representation of hESC ChIP-seq profiles at selected loci. Binding profiles for p300, H3K4me1, H3K27ac, H3K27me3 and H3K4me3 are shown for representative examples of (**a**) Class I (e.g. *OCT4/POU5F1*) and (**b-c**) tissue-specific enhancers (*AFP, CSF2, ADIPOQ, REN*) and their flanking regions. The height of the peaks corresponds to normalized fold-enrichments as calculated by QuEST (see Methods).



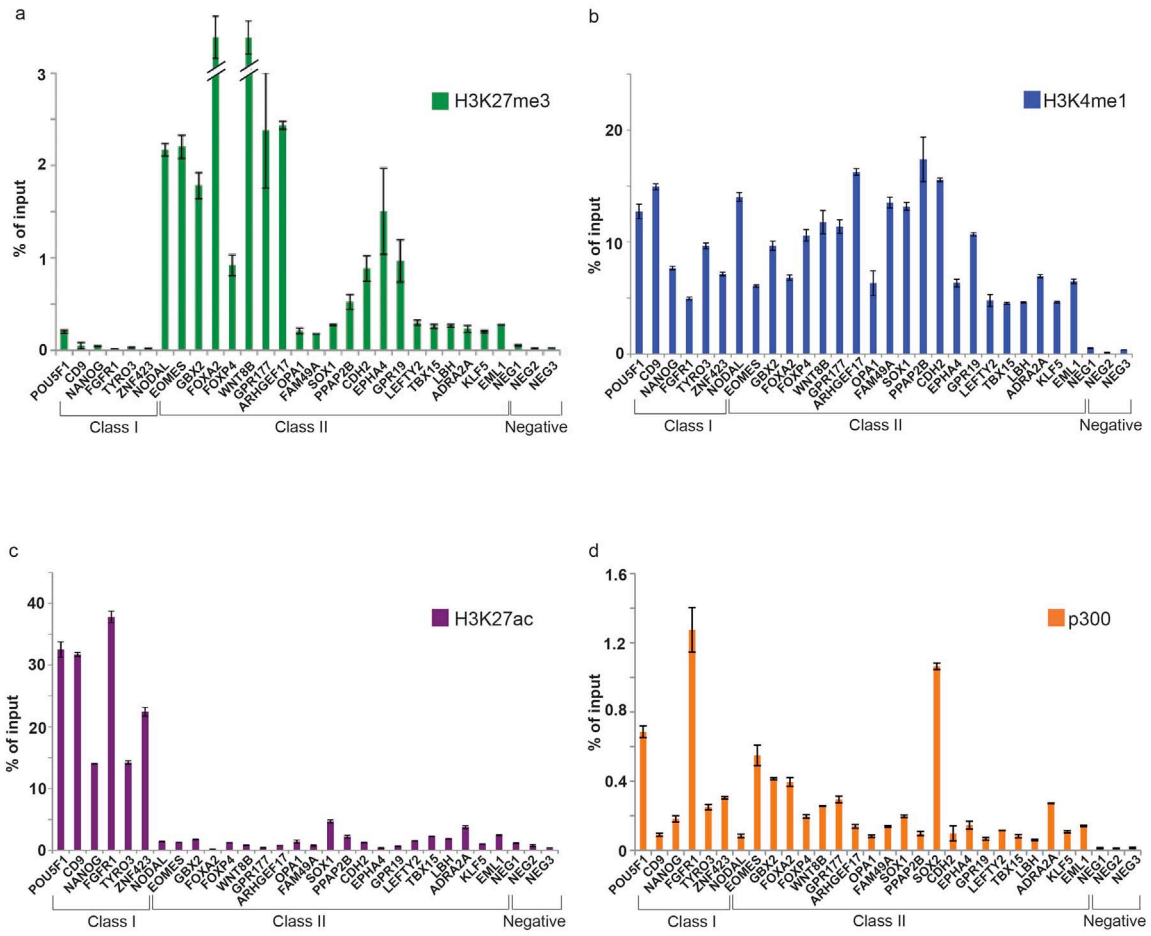
Supplementary Figure 3. Class I and Class II showed similar profiles for various enhancer features.

a, Average hESC p300 ChIP-seq signal profiles around central position of p300-bound regions for both Class I (red) and Class II (purple) elements. **b**, Average hESC BRG1 ChIP-seq signal profiles around central position of p300-bound regions for both Class I (red) and Class II (purple) elements. **c**, Average hESC FAIRE-seq signal profiles around central position of p300-bound regions for both Class I (red) and Class II (purple) elements demonstrating nucleosome depletion at Class I and Class II elements. **d**, Average hESC MRE-seq signal profiles around central position of p300-bound regions for both Class I (red) and Class II (purple) elements, demonstrating DNA hypomethylation at Class I and Class II elements.

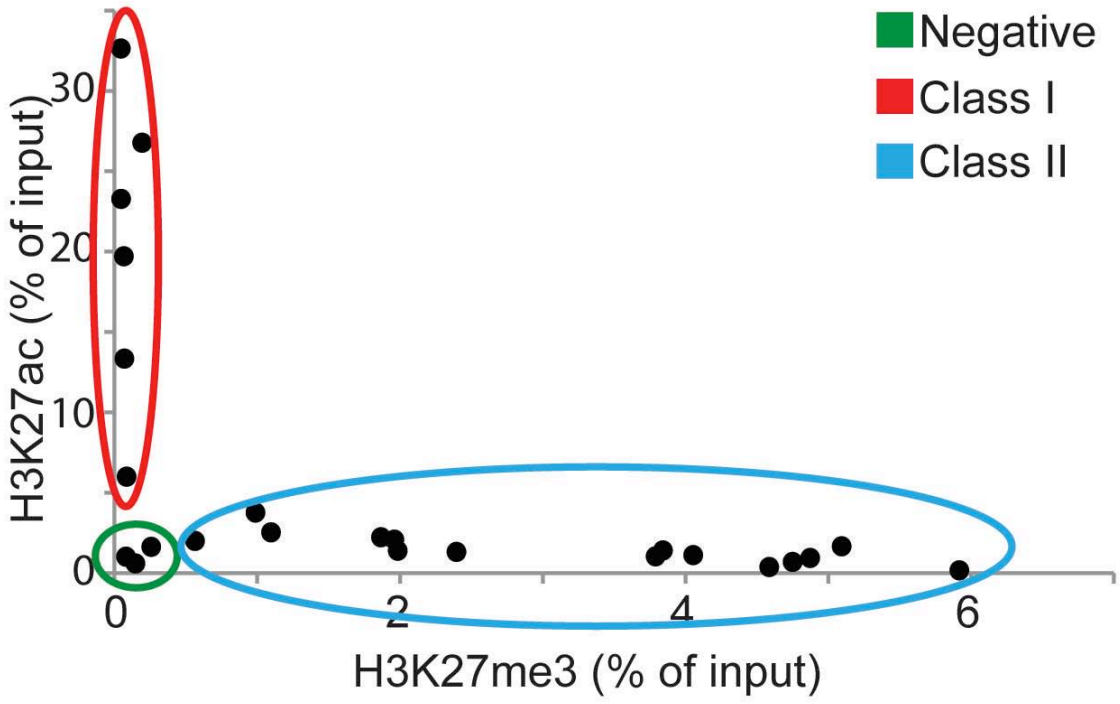


Supplementary Figure 4. ChIP-qPCR validation of ChIP-seq results in hESC. ChIP-qPCR experiments were performed from sonicated hESC chromatin using antibodies against (**a**) H3K27me3, (**b**) H3K4me1, (**c**) H3K27ac, (**d**) p300 and IgG and (**e**) SUZ12. Enrichments were calculated as % of input

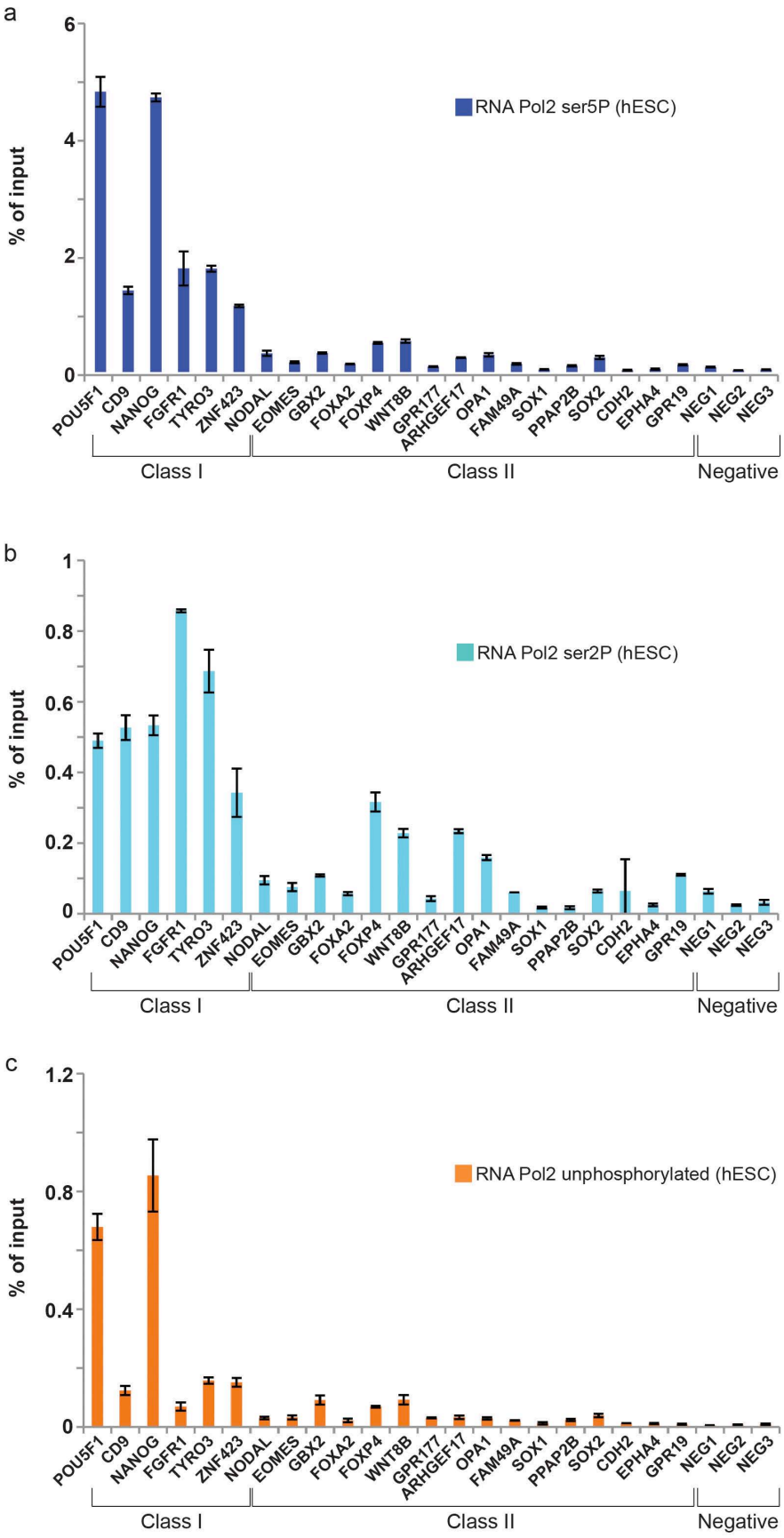
recovery for selected Class I and Class II elements and negative control regions. Error bars represent s.d. from three technical replicates. ChIPs against H3K27me3, H3K4me1, H3K27ac and p300 represent biological replicates of those samples used in ChIP-seq.



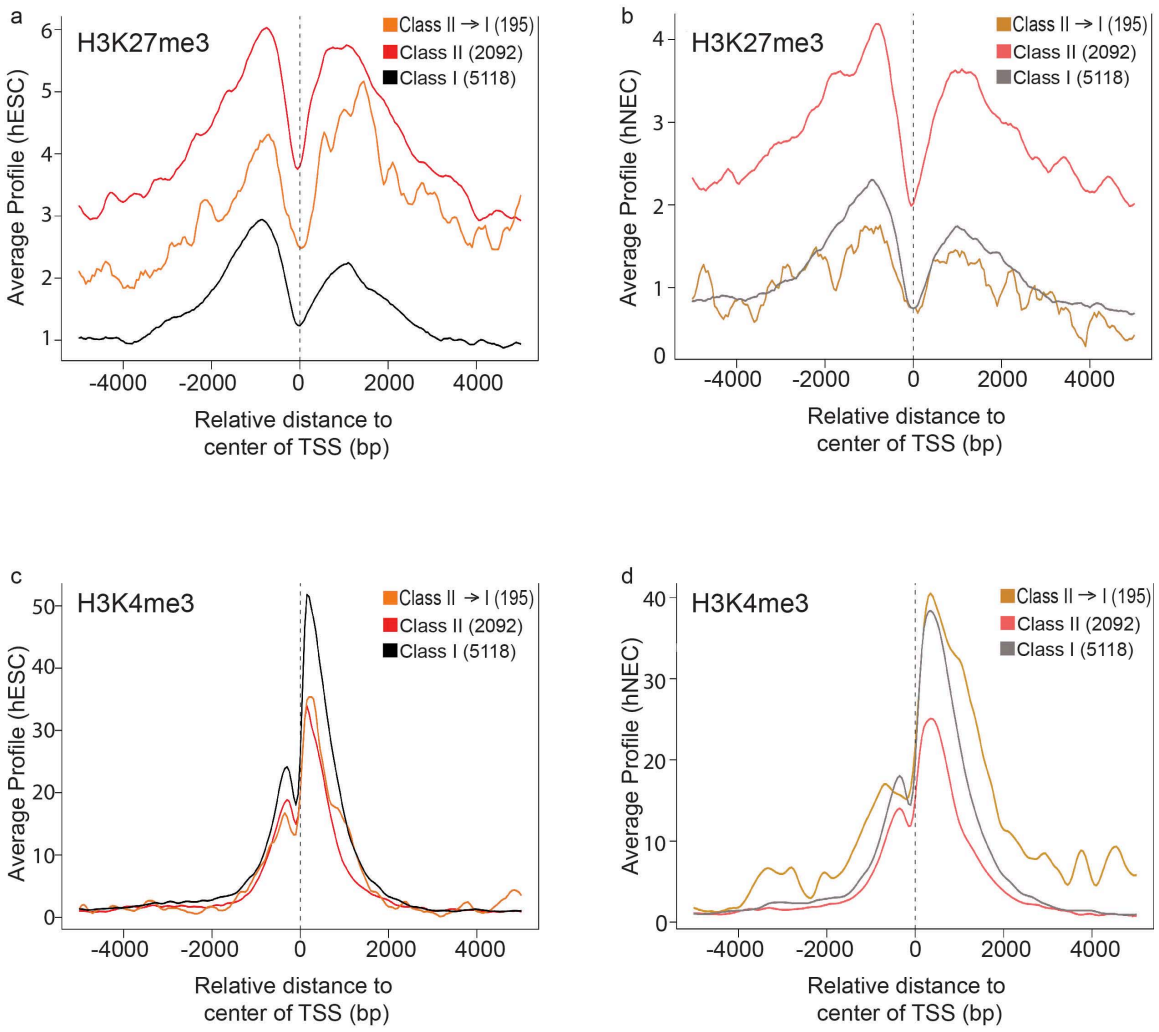
Supplementary Figure 5. Additional biological replicates of ChIP-qPCR validations in hESC. ChIP-qPCR experiments were performed as in Supplementary Figure 3 using a second set of biological ChIP replicates from sonicated hESC chromatin using antibodies against (**a**) H3K27me3, (**b**) H3K4me1, (**c**) H3K27ac and (**d**) p300. Enrichments were calculated as % of input recovery for selected Class I and Class II elements and negative control regions. Additional Class II elements were analyzed in this second biological replicates set. Error bars represent s.d. from three technical replicates.



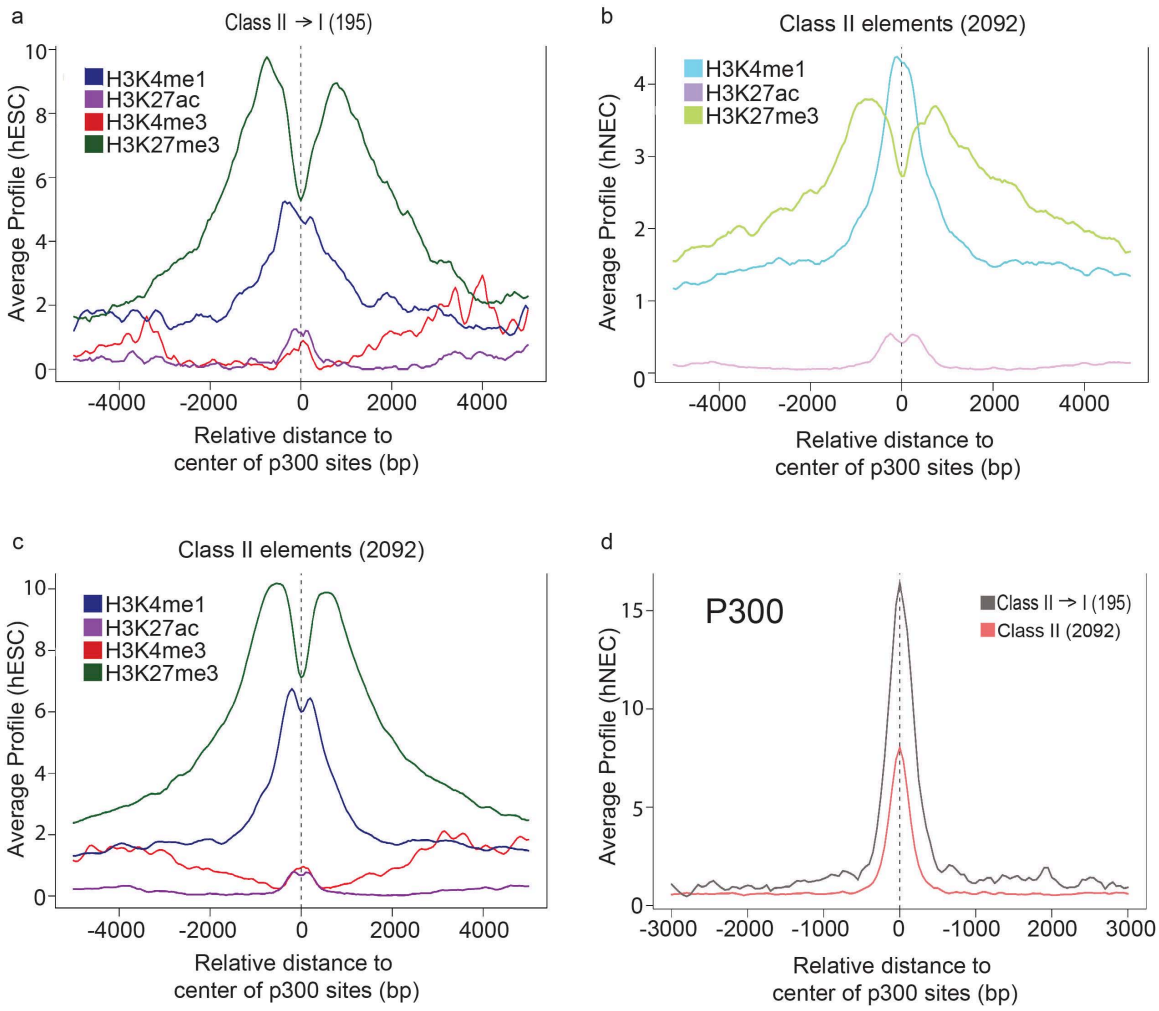
Supplementary Figure 6. H3K27me3 and H3K27ac display mutually exclusive patterns at Class I and Class II elements. H3K27me3 (x-axis) and H3K27ac (y-axis) ChIP-qPCR signals at regions interrogated in Supplementary Fig. 4a,c were plotted.


a
% of input
RNA Pol2 ser5P (hESC)
b
% of input
RNA Pol2 ser2P (hESC)
c
% of input
RNA Pol2 unphosphorylated (hESC)
POU5F1 CD9 NANOG FGFR1 TYRO3 ZNF423 NODAL EOMES GBX2 FOXA2 FOXP4 WNT8B GPR177 ARHGEF17 OPA1 FAM49A SOX1 PPAP2B SOX2 CDH2 EPHA4 GPR19 NEG1 NEG2 NEG3
Class I
Class II
Negative

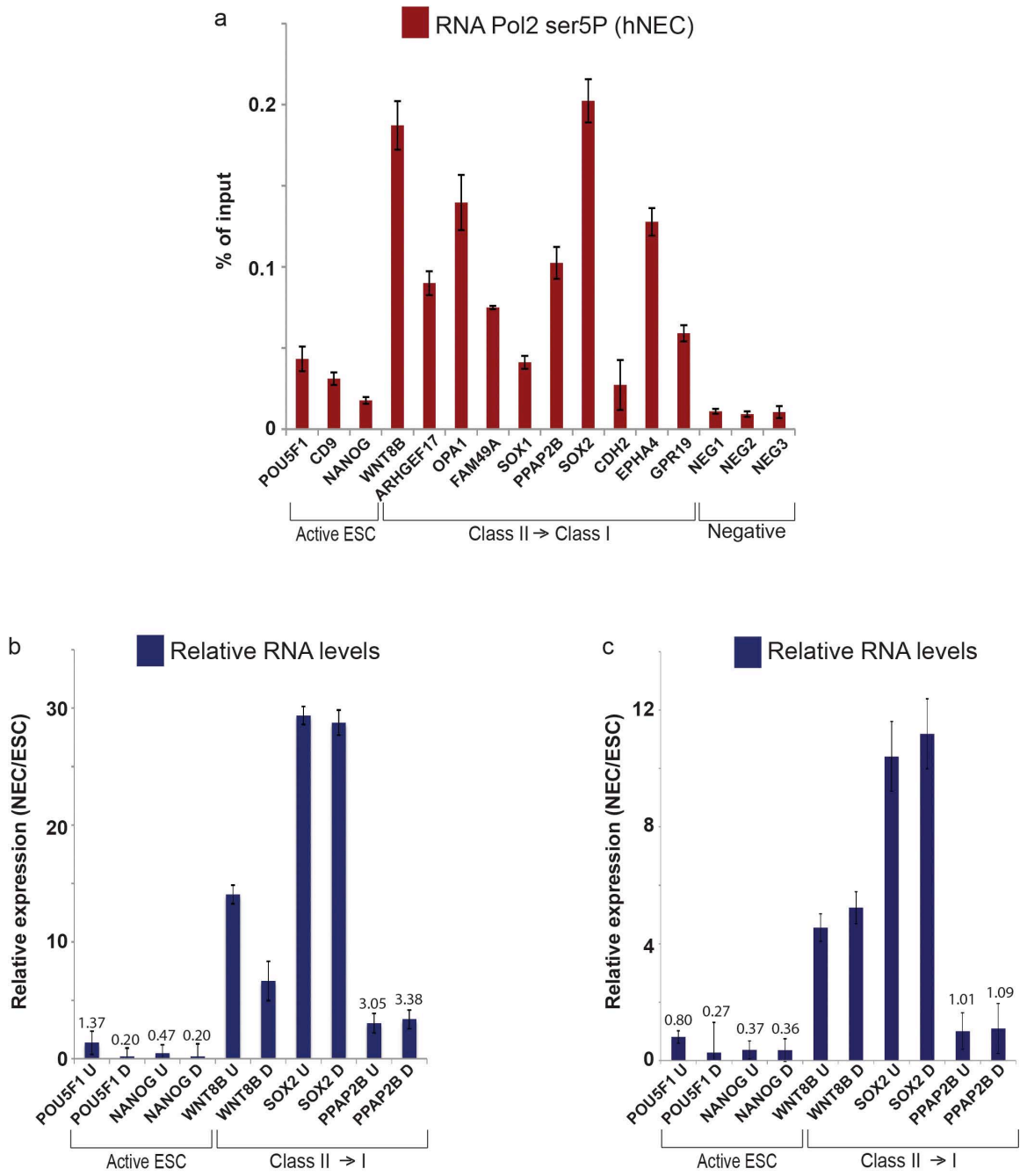
Supplementary Figure 7. RNA Pol2 binding at Class I and Class II elements in hESCs. **a-c**, ChIP-qPCR experiments were performed in hESC with antibodies against RNA Pol2 ser5P (**a**), RNA Pol2 ser2P (**b**) and unphosphorylated RNA Pol2 (**c**). Enrichments were calculated as % of input recovery for selected Class I and Class II elements and negative control regions. Error bars represent s.d. from three technical replicates.



Supplementary Figure 8. Class II elements are linked to genes with H3K27me3-marked promoters. Average H3K27me3 and H3K4me3 ChIP-seq signal profiles from hESCs (**a,c**) or hNECs (**b,d**) were generated around the TSS of genes linked to Class I , Class II and Class II→I elements, as defined in the text.

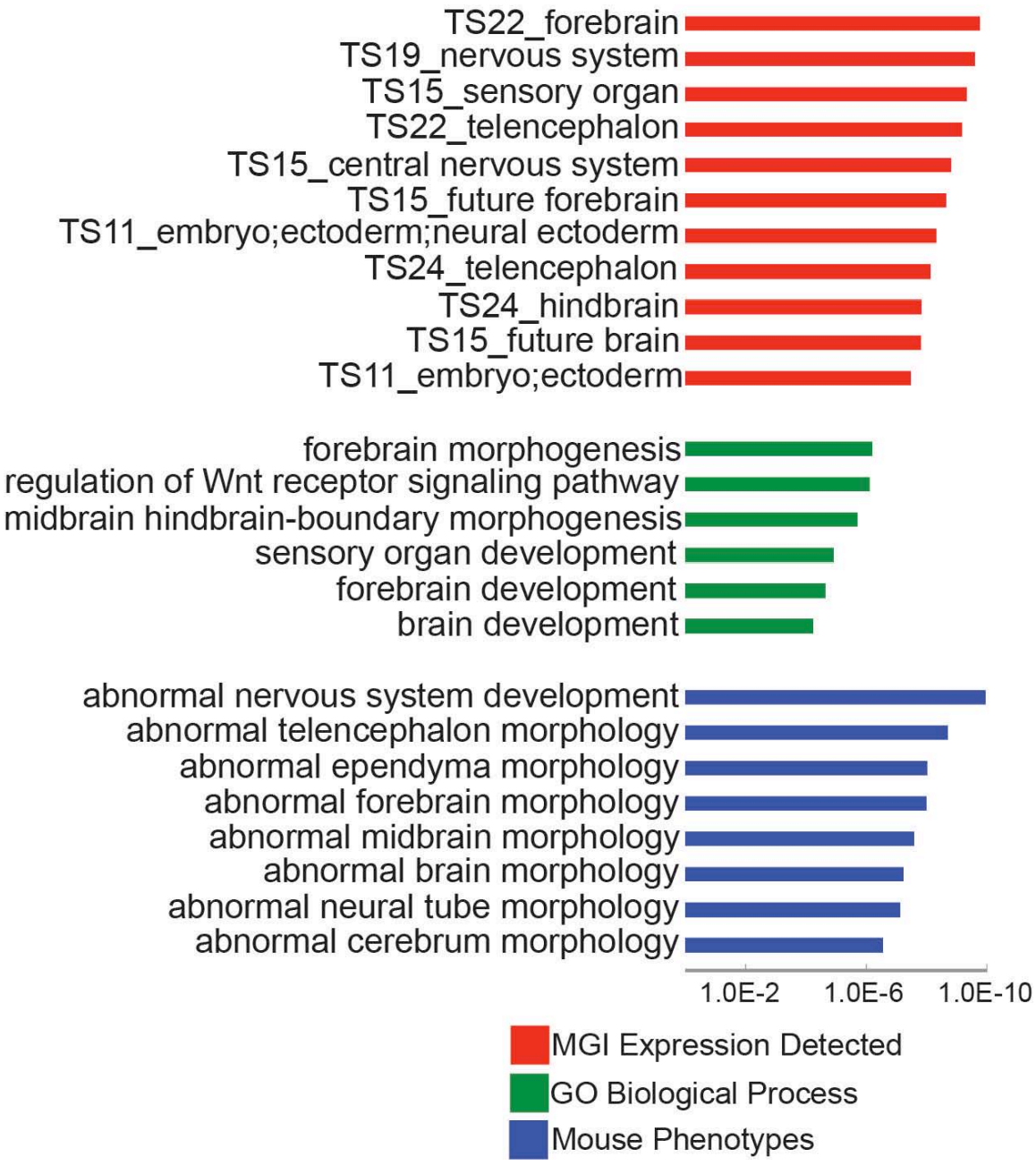


Supplementary Figure 9. Comparisons between chromatin signature profiles at Class II→I and Class II elements in hESCs and hNECs. Average ChIP-seq enrichment profiles were generated for the indicated histone modifications and p300 around the central position of Class II→I elements (e.g. those that acquired H3K27ac in hNECs) and the remaining Class II elements. **a,** hESC histone modification profiles at Class II→I elements (the corresponding profiles from hNEC are shown in Fig. 3a), **b-c,** hNEC ***(b)*** and hESC ***(c)*** histone modification profiles at Class II elements, **d,** comparison of p300 enrichment levels at Class II→I and Class II elements in hNECs.

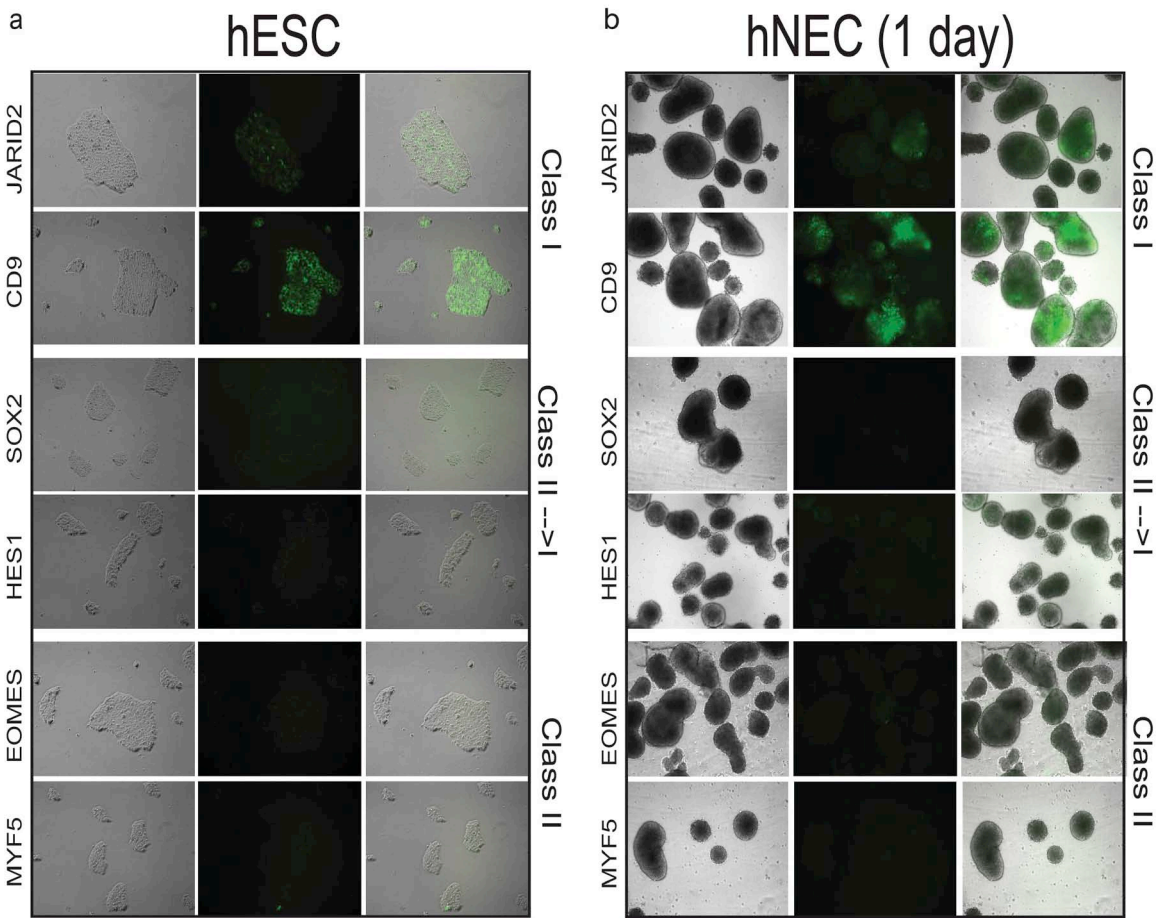


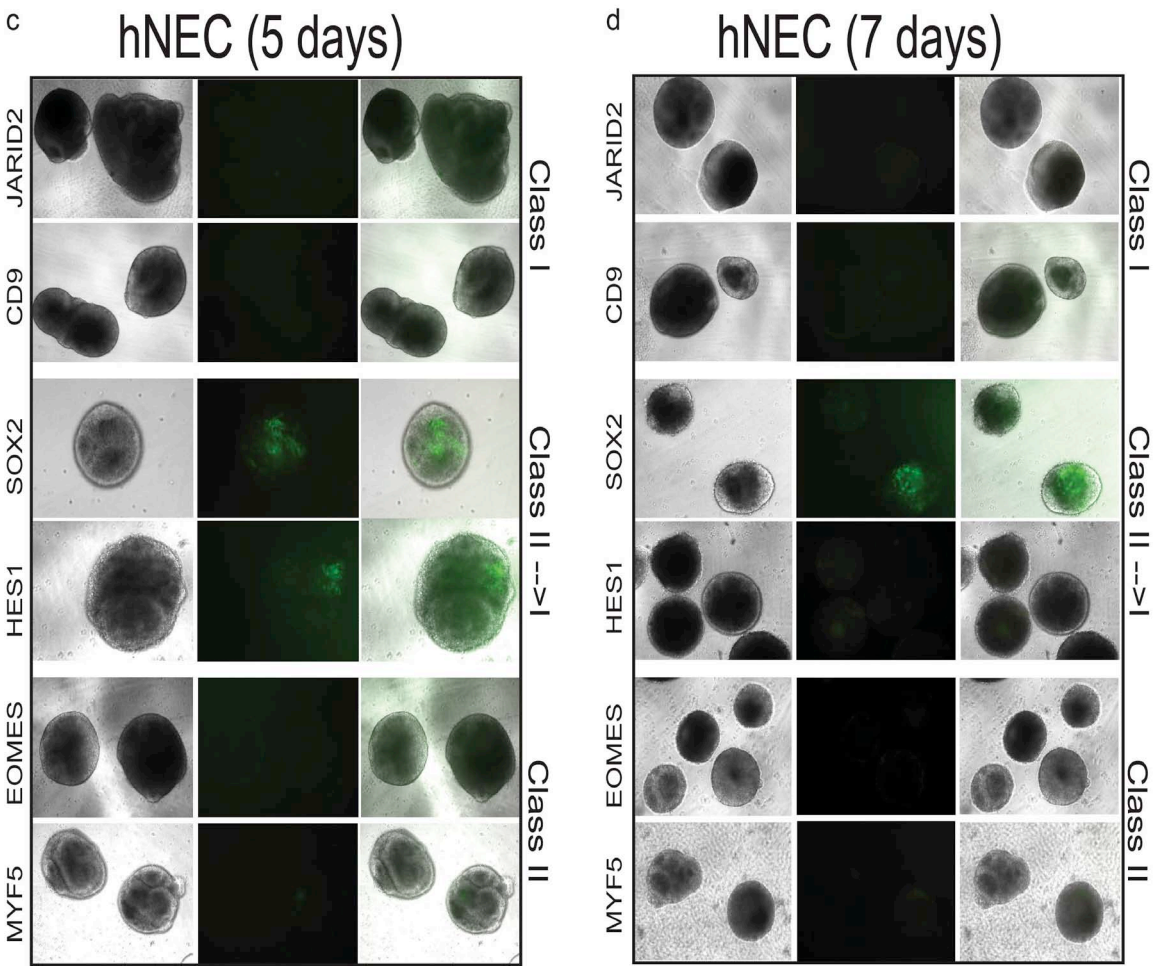
Supplementary Figure 10. Increased RNA Pol2 recruitment and bidirectional transcription at Class II→I elements in hNEC. **a**, RNA Pol2 ser5P ChIP-qPCR analysis was performed in hNEC at select Class I elements uniquely active in hESC (Active ESC) and for a subset of Class II→I elements. Enrichments were calculated as % of input recovery for selected Class I and Class II elements and negative control regions. Error bars represent s.d. from

three technical replicates. **b-c**, Bidirectional transcription from Class II→I elements. Intergenic (i.e. non-intronic) elements were selected among the enhancers described in ***a*** in order to ensure that only non-genic transcription was assayed. Transcripts originating from these elements were analyzed in by RT-qPCR in hESC and hNEC using two primer sets, corresponding to U and D regions flanking the p300 binding peak. U and D refer to upstream and downstream positions with respect to the direction of the proximal candidate target gene. The Y-axes show fold-change between hNEC and hESC transcript levels, normalized to 18S rRNA. Error bars represent s.d. from three technical replicates. All experiments were done as two biological replicates (b,c).

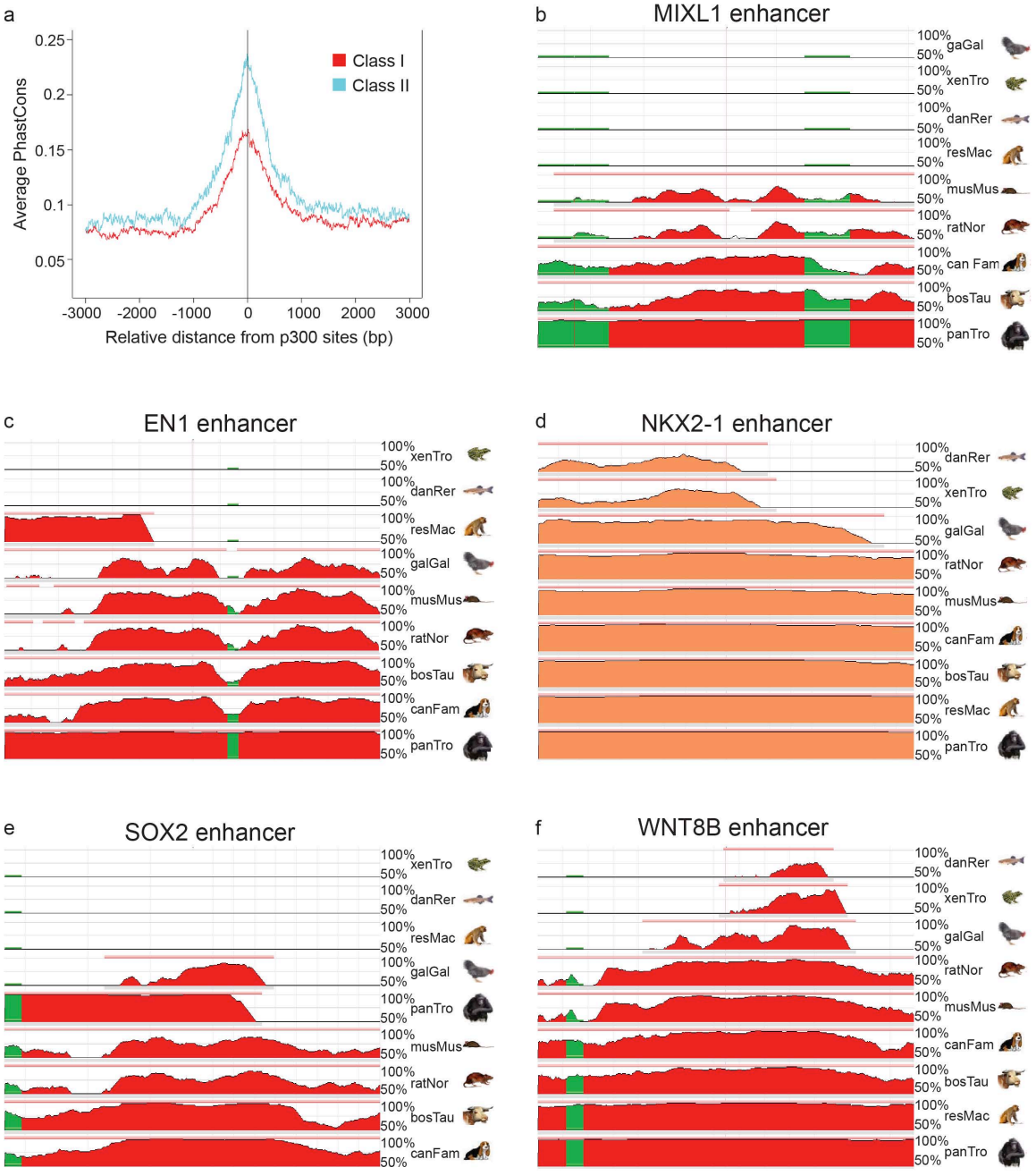


Supplementary Figure 11. Functional annotation of the Class II→I elements. Functional analysis of genes associated with Class II→I elements was performed using GREAT. Top overrepresented gene categories belonging to three different ontologies (as described in Fig. 2) are shown (see Supplementary Data 4 for additional information).


a
hESC
b
hNEC (1 day)
JARID2
CD9
SOX2
HES1
EOMES
MYF5
Class I
Class II -->I
Class II

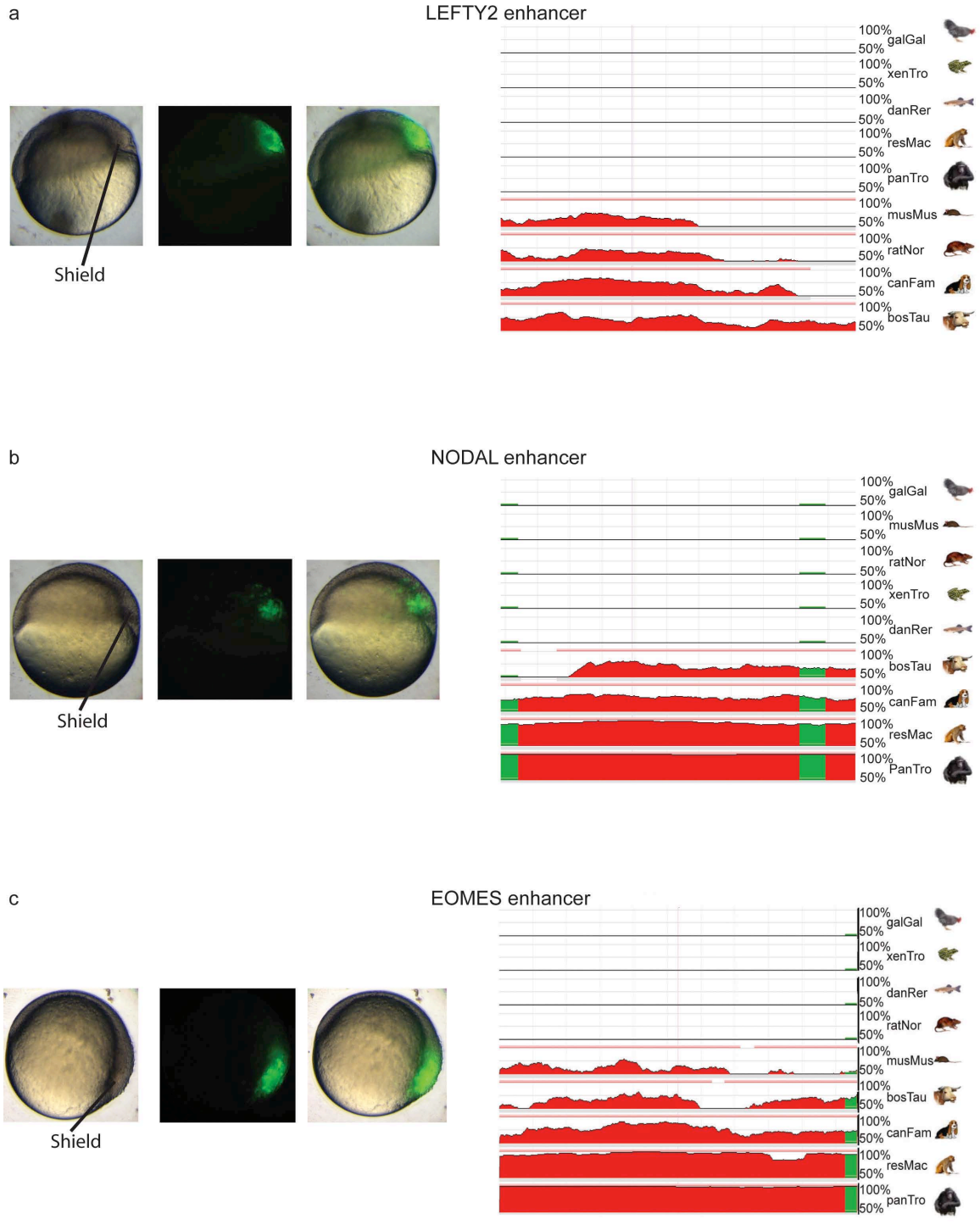


Supplementary Figure 12. *In vitro* enhancer reporter assays in hESC and during differentiation to hNEC. Select Class I (*JARID2, CD9*), Class II→I (*SOX2, HES1*) and Class II (*EOMES, MYF5*) elements were cloned in front of the minimal TK promoter into a lentiviral GFP reporter vector. hESC colonies were transduced with indicated lentiviruses and GFP fluorescence was assayed in undifferentiated hESC ***(a)*** and 1 day ***(b)***, 5 days ***(c)*** and 7 days ***(d)*** after induction to neuroepithelial spheres. For each reporter brightfield, GFP and merged images are shown, from left to right, respectively. Genomic coordinates of all elements are listed in Supplementary Table 1.



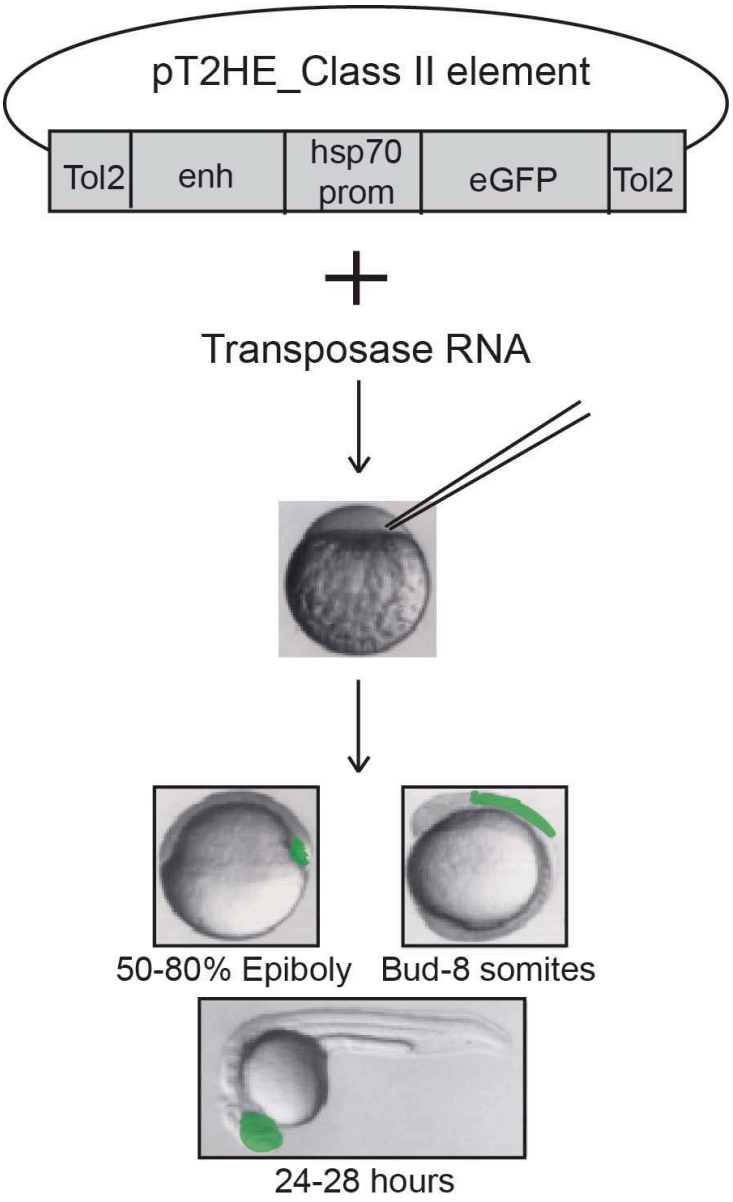
Supplementary Figure 13. Evolutionary conservation of Class II elements. **a**, Average vertebrate PhastCons profiles were generated around the central position of p300-enriched regions, both for Class I (red) and Class II (blue)

elements identified in hESC. Vertebrate PhastCons is a conservation scoring system obtained after generating multiple alignments based on the genomic sequence of 28 different vertebrate species. **b-f**, Evolutionary conservation for five of the *in vivo* assayed Class II elements. Conservation (as % of sequence identity, Y-axis) of human enhancer sequences is shown using the Evolutionary Conserved Regions (ECR, http://ecrbrowser.dcode.org/) browser for a number of vertebrates: chicken (galGal), frog (xenTro), zebrafish (danRer), rat (ratNor), mouse (musMus), cow (bosTau), dog (canFam), macaque (resMac) and chimpanzee (panTro). Red, salmon and green colors correspond to intergenic, intronic and repetitive genomic regions, respectively.

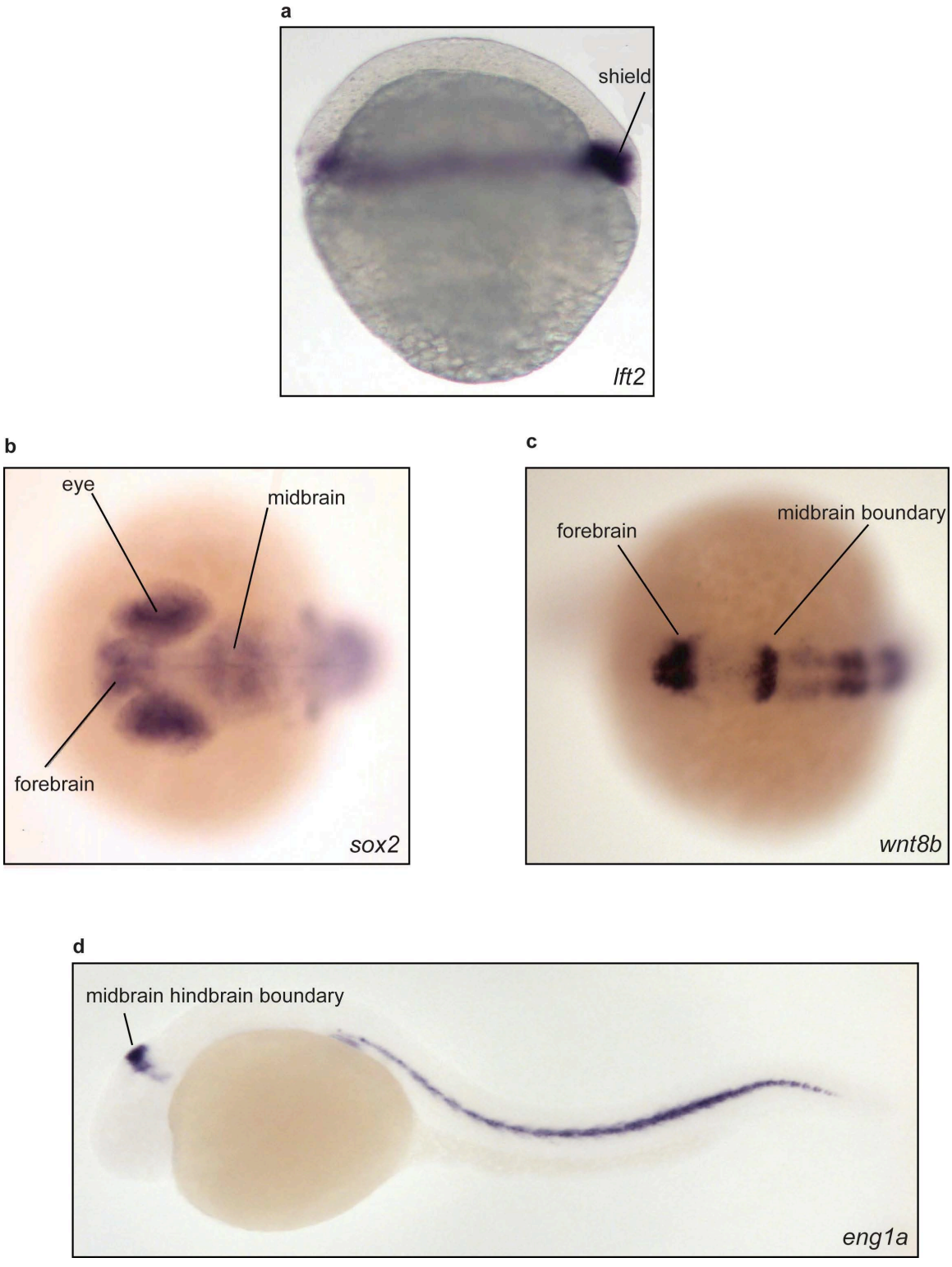


Supplementary Figure 14. Human Class II elements can regulate gastrulation-specific gene expression in zebrafish embryos. Left panels: representative brightfield, GFP fluorescence, and merged images of the shield

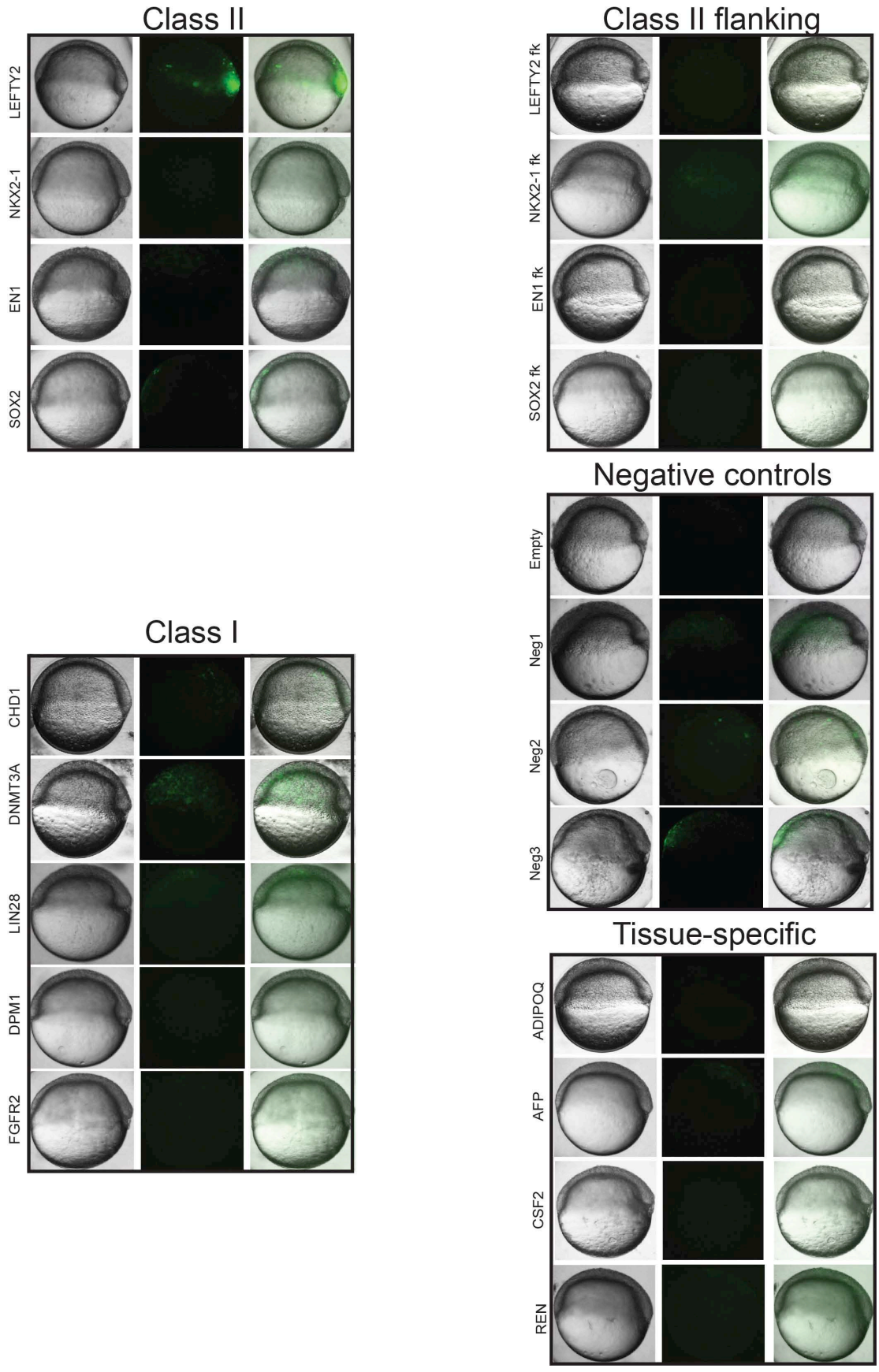
stage (6 hpf) zebrafish embryos co-injected at 1-cell stage with transposase and enhancer reporter constructs for human (**a**) *LEFTY2*, (**b**) *NODAL* and (**c**) EOMES Class II elements. Lateral (animal pole to the top) views are shown. Right panels, evolutionary conservation (as % of sequence identity, Y-axis) of human *LEFTY2*, *NODAL* and *EOMES* class II element sequences is shown using the Evolutionary Conserved Regions (ECR, http://ecrbrowser.dcode.org/) browser for a number of vertebrates: chicken (galGal), frog (xenTro), zebrafish (danRer), rat (ratNor), mouse (musMus), cow (bosTau), dog (canFam), macaque (resMac) and chimpanzee (panTro). Red and green colors correspond to intergenic and repetitive genomic regions, respectively.



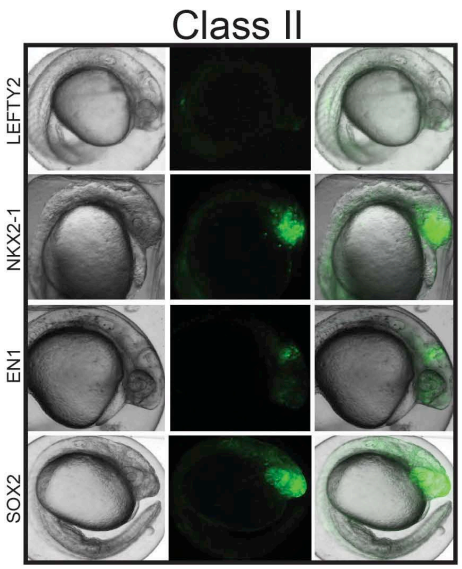
Supplementary Figure 15. Schematic representation of the enhancer reporter assay in zebrafish. DNA fragments corresponding to select human Class II elements were cloned upstream of the hsp70 minimal promoter driving eGFP expression into a vector containing Tol2 transposase sequences. The resulting constructs (pT2HE_Class II elements) were injected, together with transposase RNA, into 1-cell stage zebrafish embryos.

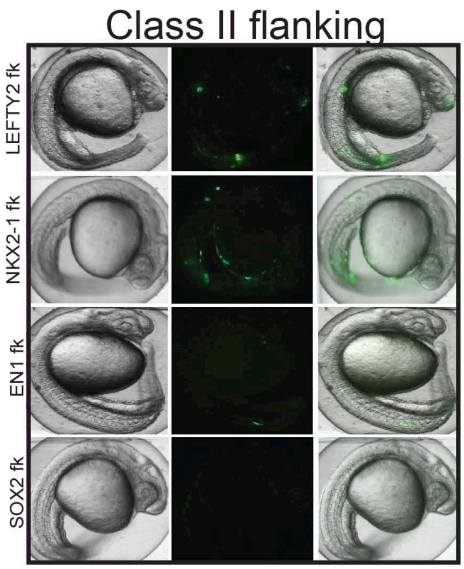

a
shield
lft2
b
eye
midbrain
forebrain
sox2
c
forebrain
midbrain boundary
wnt8b
d
midbrain hindbrain boundary
eng1a

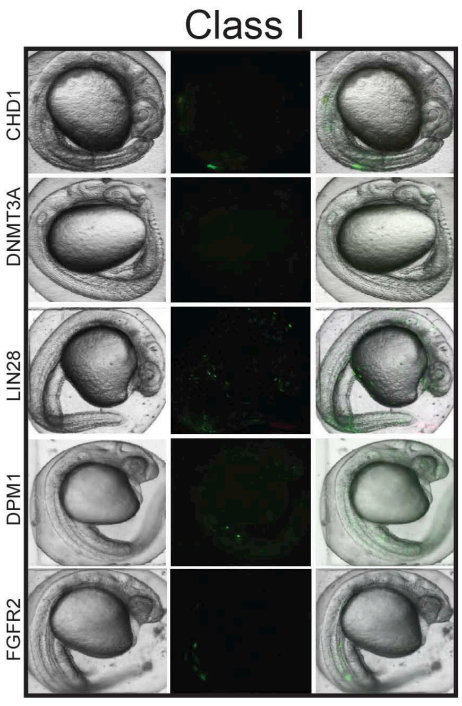
Supplementary Figure 16. Expression patterns of zebrafish homologs of Class II element putative target genes. RNA *in situ* hybridization data was obtained from the Zebrafish Model Organism Database (ZFIN, http://zfin.org) for zebrafish homologs of a subset of the Class II element putative target genes (e.g. genes located proximally to the identified elements). Corresponding elements were tested in the *in vivo* enhancer assay as shown in Figure 4. Additional RNA *in situ* data for these and zebrafish homologs of the remaining genes proximal to the assayed Class II elements (i.e. *EOMES, NODAL, NKX2-1, MIXL1*) can be found in the ZFIN database. (**a**) *lft2*, zebrafish homolog of human *LEFTY2,* shield-stage, lateral view (ZFIN ID: ZDB-IMAGE-030925-334), (**b**) *sox2,* zebrafish homolog of human *SOX2* Prim-5 (24 hpf) stage, dorsal anterior view (ZFIN ID: ZDB-IMAGE-030925-260), (**c**) *wnt8b,* zebrafish homolog of human *WNT8B,* 14-19 somite stage (16-19 hpf), dorsal anterior view (ZFIN ID: ZDB-IMAGE-060130-306), (**d**) *eng1a,* zebrafish homolog of human *EN1*, Prim-5 (24hpf) stage, lateral view (ZFIN ID: ZDB-IMAGE-050208-1445). Only *EOMES* and *MIXL1* gene homologs showed expression patterns significantly distinct from those driven by the proximal human Class II elements. Interestingly, however, expression patterns observed for these two genes in other vertebrates and specifically in mammals [28,33,35], correspond to zebrafish embryonic structures in which we detected enhancer-driven expression.

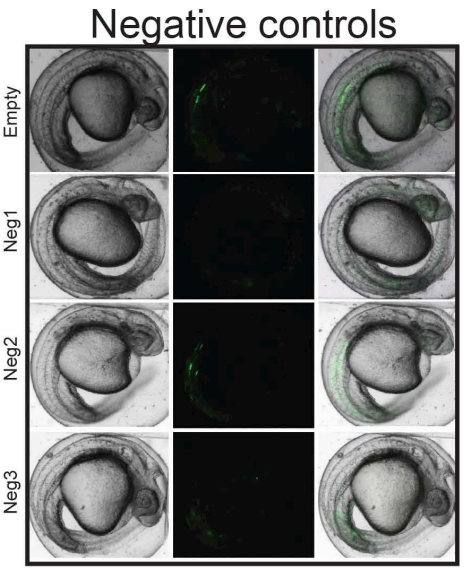

Class II
LEFTY2
NKX2-1
EN1
SOX2
Class II flanking
LEFTY2 fk
NKX2-1 fk
EN1 fk
SOX2 fk
Negative controls
Empty
Neg1
Neg2
Neg3
Class I
CHD1
DNMT3A
LIN28
DPM1
FGFR2
Tissue-specific
ADIPOQ
AFP
CSF2
REN

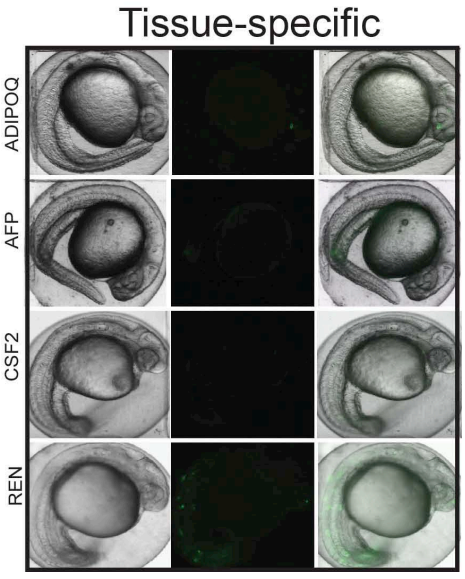
Supplementary Figure 17. GFP fluorescence driven by select Class II elements and control genomic regions in zebrafish embryos at 6 hpf. A select set of enhancer reporters representing Class II elements, Class I elements and different categories of genomic control regions, indicated on the top of each section (e.g. regions flanking the assayed Class II elements, negative regions represented by the three randomly selected non-coding genomic sequences and empty vector control, as well as previously characterized adult tissue-specific enhancers) were assayed for GFP fluorescence at the shield stage (6 hpf). Representative images are shown in lateral view with shield to the right. For each element tested, brightfield, GFP and merged images are shown, from left to right, respectively. Genomic coordinates of all regions are listed in Tables 3 and 4.


Class II
LEFTY2
NKX2-1
EN1
SOX2

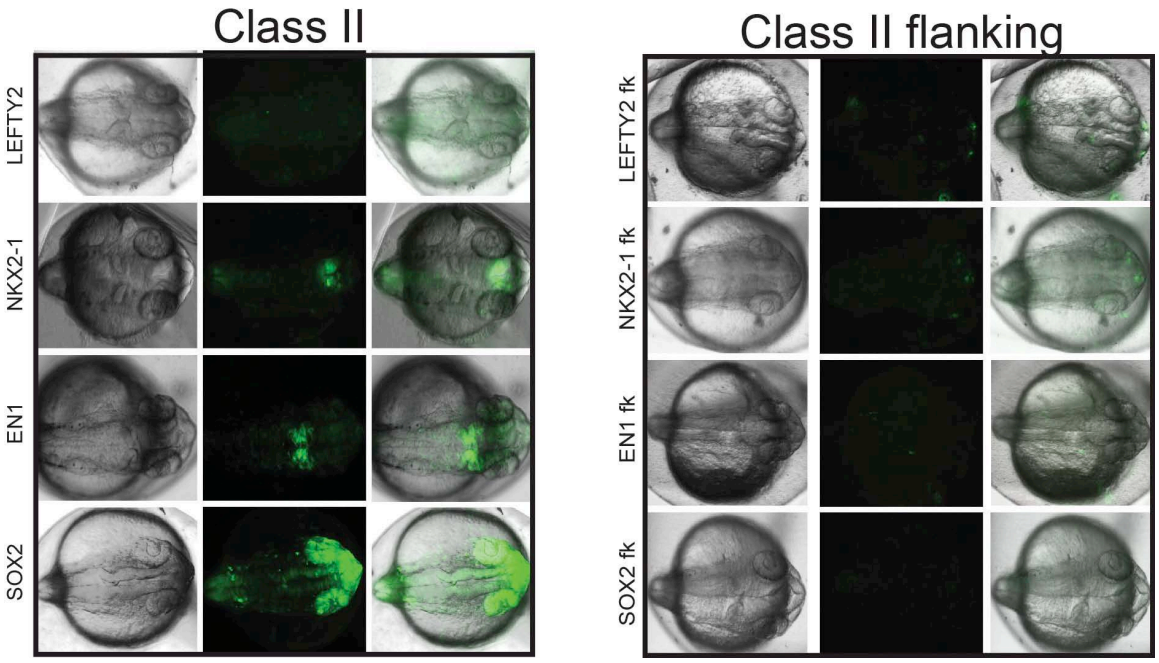

Class II flanking
LEFTY2 fk
NKX2-1 fk
EN1 fk
SOX2 fk


Class I
CHD1
DNMT3A
LIN28
DPM1
FGFR2

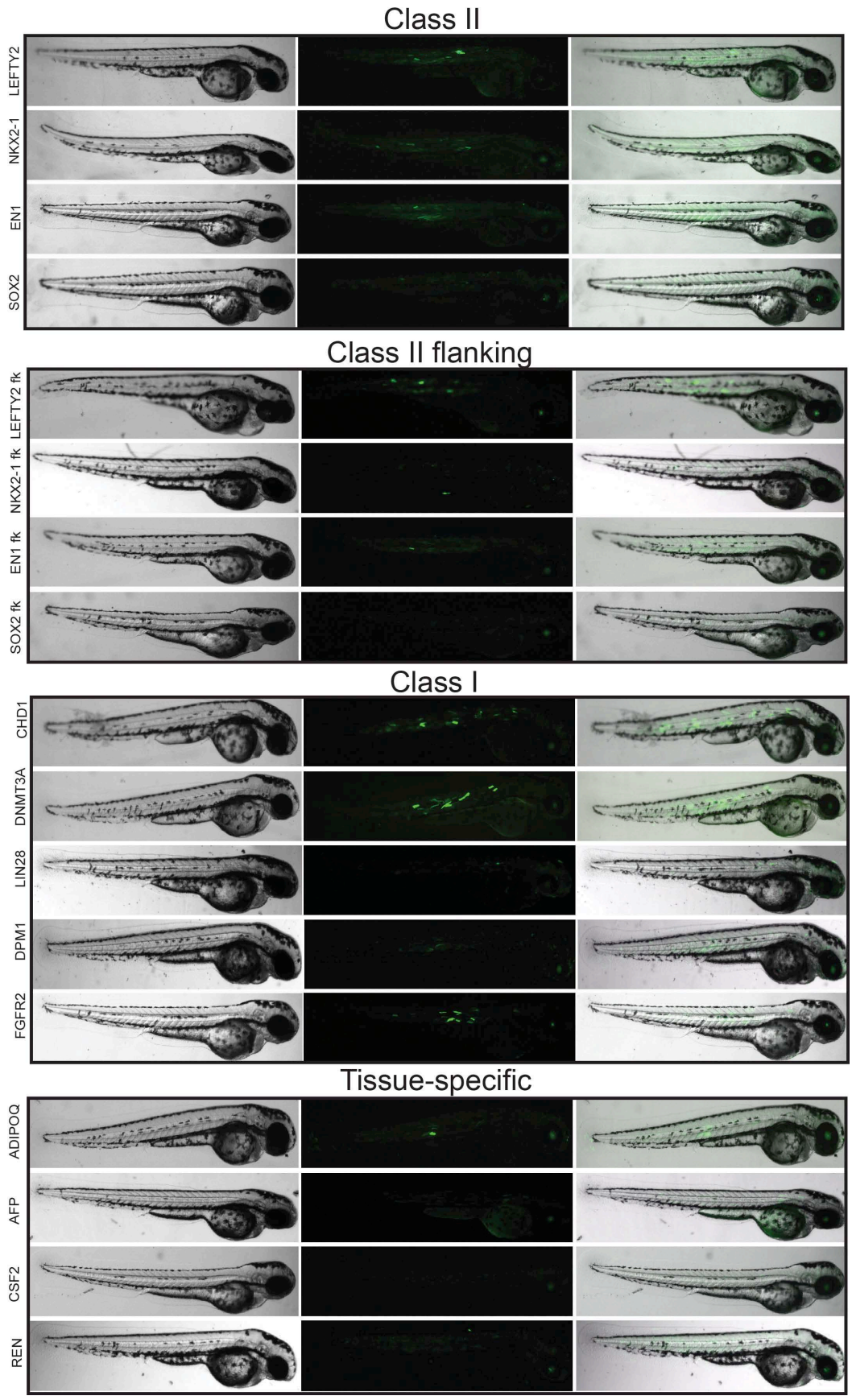

Negative controls
Empty
Neg1
Neg2
Neg3


Tissue-specific
ADIPOQ
AFP
CSF2
REN

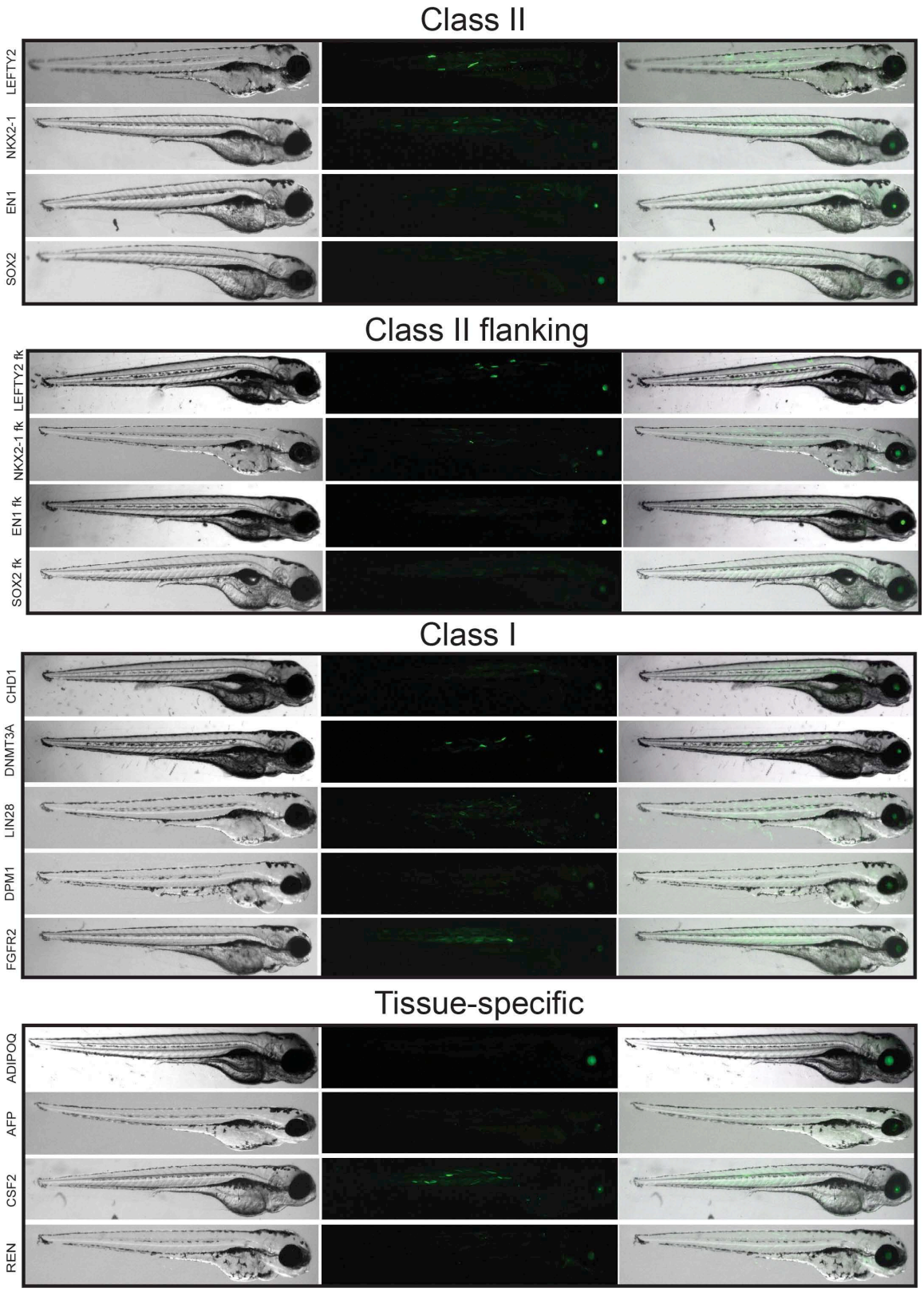
Supplementary Figure 18. GFP fluorescence driven by select Class II elements and control genomic regions in zebrafish embryos at 24 hpf. A select set of enhancer reporters representing Class II elements, Class I elements and different categories of genomic control regions, indicated on the top of each section (e.g. regions flanking the assayed Class II elements, negative regions represented by the three randomly selected non-coding genomic sequences and empty vector control, and previously characterized adult tissue-specific enhancers) were assayed for GFP fluorescence at 24 hpf. Representative images are shown in lateral view without removing the chorion. For each element tested, brightfield, GFP and merged images are shown, from left to right, respectively. Genomic coordinates of all regions are listed in Tables 3 and 4.



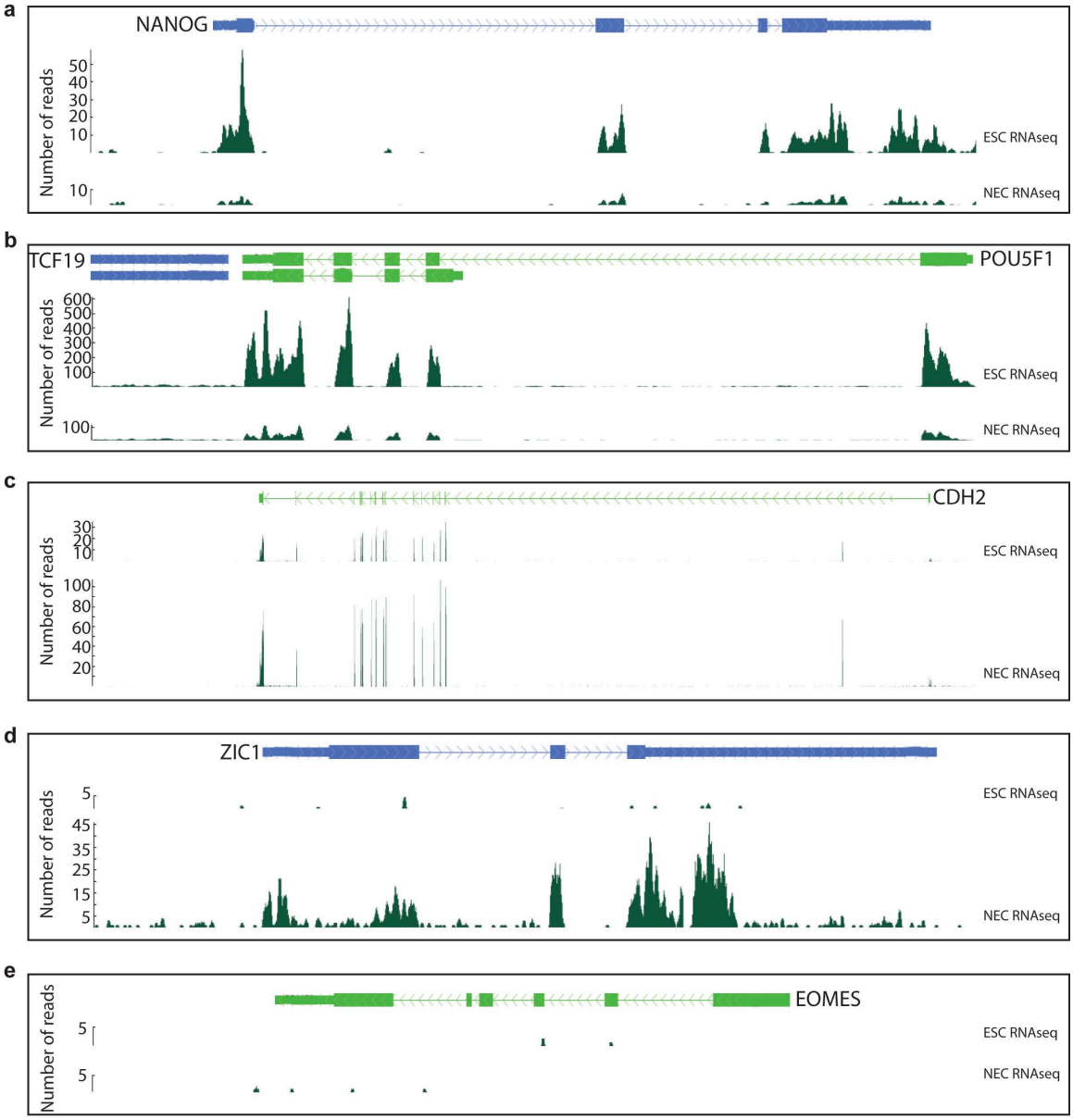
Supplementary Figure 19. GFP patterns for select Class II elements and flanking regions after 24 hpf in zebrafish embryos. For each element tested, brightfield, GFP and merged images are shown, from left to right. All embryos were imaged after 24 hpf and are shown in dorsal anterior view, with anterior to the right.


Class II
LEFTY2
NKX2-1
EN1
SOX2
Class II flanking
LEFTY2 fk
NKX2-1 fk
EN1 fk
SOX2 fk
Class I
CHD1
DNMT3A
LIN28
DPM1
FGFR2
Tissue-specific
ADIPOQ
AFP
CSF2
REN

Supplementary Figure 20. GFP fluorescence driven by select Class II elements and control genomic regions in zebrafish embryos at 3 dpf. A select set of enhancer reporters representing Class II elements, Class I elements and different categories of genomic control regions, indicated on the top of each section (e.g. regions flanking the assayed Class II elements, and previously characterized adult tissue-specific enhancers) were assayed for GFP fluorescence at 3 dpf. Representative images are shown in lateral view with anterior to the right. For each element tested, brightfield, GFP and merged images are shown, from left to right, respectively. Genomic coordinates of all regions are listed in Tables 3 and 4.


Class II
LEFTY2
NKX2-1
EN1
SOX2
Class II flanking
LEFTY2 fk
NKX2-1 fk
EN1 fk
SOX2 fk
Class I
CHD1
DNMT3A
LIN28
DPM1
FGFR2
Tissue-specific
ADIPOQ
AFP
CSF2
REN

Supplementary Figure 21. GFP fluorescence driven by select Class II elements and control genomic regions in zebrafish embryos at 5 dpf. A select set of enhancer reporters representing Class II elements, Class I elements and different categories of genomic control regions, indicated on the top of each section (e.g. regions flanking the assayed Class II elements, and previously characterized adult tissue-specific enhancers) were assayed for GFP fluorescence at 5 dpf. Representative images are shown in lateral view with anterior to the right. For each element tested, brightfield, GFP and merged images are shown, from left to right, respectively. Genomic coordinates of all regions are listed in Tables 3 and 4.



Supplementary Figure 22. Examples of hESC and hNEC RNA-seq data. Genome browser representation of RNA-seq results from hESC and hNEC is shown as number of reads (both sense and antisense) at selected loci representing genes active in hESC (**a**, **b**), active in hNEC (**c**, **d**) and inactive in both cell types (**e**). RNA-seq analysis and browser pictures were obtained using DNAnexus (https://dnanexus.com).

Supplementary Table 1. Distal elements tested for *in vitro* enhancer activity in gene reporter assays.

HG18 coordinates	Associated Gene	Class
chr12:6,194,198-6,195,060	*CD9*	Class I
chr6:15,316,353-15,317,220	*JARID2*	Class I
chr3:182956946-182957845	*SOX2*	Class II →I
chr3:194,975,593-194,976,914	*HES1*	Class II →I
chr3:27744727-27745788	*EOMES*	Class II
chr12:79,632,491-79,633,408	*MYF5*	Class II

For each of the distal elements, from left to right, we present: (i) the genomic coordinates according to hg18 human genome assembly; (ii) the closest gene to each of the elements and (iii) their class as defined in methods.

Supplementary Table 2. Enhancer activity of hESC Class II elements in mouse embryos at E11.5 developmental stage.

HG18 coordinates	Vista Enhacer ID	Closest gene	Expression pattern
chr8:95,069,303-95,070,373	hs26	IRX6	Eye, midbrain, neural tube
chr2:176765577-176766917	hs246	HOXD1	Hindbrain
chr6:41631202-41631655	hs281	FOXP4	Forebrain, midbrain
chr13:27293961-27295536	hs532	GSX1	Forebrain, hindbrian, midbrain
chr8:53329283-53330389	hs698	ST18	Dorsal root ganglion, forebrain, hindbrain, midbrain, neural tube, trigeminal V
chr5:87118768-87120508	hs853	CCNH	Forebrain
chr10:102234832-102236324	hs1006	WNT8B	Forebrain, hindbrain, midbrain
chr11:16379845-16382613	hs1301	SOX6	Forebrain
chr18:23917861-23922854	hs1634	CDH2	Branchial arch, cranial nerve

Nine hESC Class II elements overlap with enhancers displaying specific activity during mouse development at E11.5 stage, as reported in the VISTA enhancer browser (http://enhancer.lbl.gov/). For each of the enhancers, from left to right, we present: (i) the hESC Class II element genomic coordinates according to hg18 human genome assembly; (ii) the VISTA enhancer IDs for those previously tested enhancers overlapping with Class II elements; (iii) the closest gene to each of the elements and (iv) the mouse expression patterns reported for each of the VISTA enhancers. Red color indicates those Class II elements that became Class II→I during neuroectodermal differentiation.

Supplementary Table 3. Summary of zebrafish reporter assays for an interrogated subset of hESC Class II elements.

HG18 coordinates	Associated gene	Stage	Anatomical structure
chr10:71884931-71885622	*NODAL*	50% epiboly/Shield	Shield
chr3:27744727-27745788	*EOMES*	50% epiboly/Shield	Shield
chr1:224199839-224200532	*LEFTY2*	50% epiboly/Shield	Shield, Germ ring
chr3:182956946-182957845	*SOX2*	24-28 hpf	Eyes, forebrain
chr14:36043730-36044174	*NKX2-1*	24-28 hpf	Ventral diencephalon (hypothalamus)
chr2:119281096-119282486	*EN1*	24-28 hpf	Midbrain hindbrain boundary
chr10:102234833-102236324	*WNT8B*	24-28 hpf	Forebrain, midbrain, hindbrain
chr1:224464748-224465456	*MIXL1*	24-28 hpf	Tail muscle, notochord
chr2:236732477-236733730	*GBX2*	NA	NA

Nine hESC Class II elements were used in zebrafish enhancer reporter assays and their expression patterns were monitored during the first 24-28 hours post fertilization (hpf). From left to right, we present: (i) the genomic coordinates of the tested elements according to the hg18 human genome assembly; (ii) the human genes associated with each element; (iii) the zebrafish developmental stage when a specific expression pattern was detected and (iv) the anatomical structure where such expression pattern occurred.

Supplementary Table 4. Summary of various types of control genomic regions tested in zebrafish reporter assays.

HG18 coordinates	Name	Class
chr17:2,981,663-2,982,262	*Neg1*	Negative
chr13:64,836,733-64,837,399	*Neg2*	Negative
chr5:92,705,310-92,706,142	*Neg3*	Negative
chr10:123,329,389-123,330,721	*FGFR2*	Class I
chr16:67,354,576-67,355,666	*CHD1*	Class I
chr1:26,617,600-26,618,925	*LIN28*	Class I
chr2:25,335,403-25,336,755	*DNMT3A*	Class I
chr20:49,041,751-49,043,264	*DPM1*	Class I
chr1:224,207,906-224,208,676	*LEFTY2 fk*	Class II flanking
chr2:119,293,755-119,294,799	*EN1 fk*	Class II flanking
chr14:36,039,385-36,040,467	*NKX2-1 fk*	Class II flanking
chr3:182,949,998-182,950,923	*SOX2 fk*	Class II flanking
chr5:131,434,040-131,435,039	*CSF2*	Tissue-specific
chr4:74,516,483-74,517,522	*AFP*	Tissue-specific
chr3:188,040,022-188,041,091	*ADIPOQ*	Tissue-specific
chr1:202,412,618-202,413,652	*REN*	Tissue-specific

For each of the distal elements, from left to right , we present: (i) the genomic coordinates according to hg18 human genome assembly; (ii) the name we assigned to each element, typically based on the closest gene and (iii) their class as defined in methods.